IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant:Ramnarayan *et al.*

Serial No.: 09/709,905
Confirmation No.:3606
Customore No. 24961

Filed:November 10, 2000

For:*USE OF COMPUTATIONALLY DERIVED PROTEIN STRUCTURES OF GENETIC POLYMORPHISMS IN PHARMACOGENOMICS FOR DRUG DESIGN AND CLINICAL APPLICATIONS*

Art Unit:1631

Examiner:Brusca, J.

MARKED-UP PARAGRAPHS AND CLAIMS (37 CFR §1.121)

IN THE SPECIFICATION:

Please amend the specification as follows:

Please amend the paragraph on page 3, line 22-31, as follows:

Structural changes that arise as a result of genetic polymorphisms are not of unlimited variety, since 3-D structure impacts upon function. A knowledge of the repertoire of the fine differences among generally similar 3-D structures of particular proteins will permit design of drugs that bind to [the] most polymorphisms, drugs that induce the fewest side-effects, and drugs that are more effective against infectious agents. Knowledge of these structures ultimately will permit patient-specific or subpopulation-specific, such as ethnic, age, or gender groups, design or selection of drugs.

Please amend the paragraph on page 7, lines 8-30, as follows:

A computer-based method for identifying compensatory mutations in a target protein is also provided. The method involves obtaining the amino acid sequence of a target protein containing multiple amino acid mutations that is expressed in a patient, where the structure of a form of the target protein that

responds to a particular drug, including the active site, has been structurally characterized; generating a 3-D structural model of the mutated protein; comparing the structure of the mutated protein with the form of the protein that responds to the drug to identify structural differences and/or similarities arising from the mutations; comparing the biological activities of the drug against the mutated protein and the form of the protein that responds to the drug to determine the effects of the mutations on drug response; and identifying the mutations in the protein that affect biological activity based on the comparisons. The target [biolmolecules] biomolecules can also be used in a method referred to herein as computational phenotyping to predict drug sensitivity or resistance for a given genotype. These computer-based method for identifying phenotypes *in silico* are provided. The methods involve obtaining from a patient/specimen, such as a body fluid or tissue sample, including blood, cerebral spinal fluid, urine, saliva, sweat and tissue samples, the amino acid sequence of a target protein; generating a 3-D structural model of the target protein; performing protein-drug binding analyses; and predicting drug sensitivity or resistance based on the protein-drug binding analyses.

Please amend the paragraph on page 8, line 24-31, as follows:

The databases can also be used for identification of invariant residues and regions of a target [biomoleucle] biomolecule, such as an HIV protease or reverse transcriptase. The identified invariant regions are then used to computationally screen compounds, preferably small molecules by assessing binding interactions. The compounds so-identified serve as candidates for drugs that will be effective for a larger [proporation] proportion of a population or against a broader range of variants of a pathogen, where the target protein is from a [pathogens] pathogen.

Please amend the paragraph on page 12, lines 1-7, as follows:

As used herein, structural variant proteins refer the variety of 3-D molecular structures or models thereof that result from the polymorphisms. These variants typically arise from transcription and translation of genes containing genetic polymorphisms, but also include [diffentially] differentially [glyosylated]glycosylated or otherwise post-translationally modified variants that potentially exhibit differential interactions with drugs and drug candidates.

Please amend the paragraph on page 12, lines 19-25, as follows:

As used herein, structure-based drug design refers to computer-based methods in which 3-D coordinates for molecular structures are used to identify potential drugs that can interact with a biological receptor. Examples of such methods include, but are not limited to, searching of small molecule libraries or databases, conformational searching of a ligand within an active site of [identify] identified biologically active conformations or computational docking methods.

Please replace the paragraph on page 13, lines 1-16, with the following:

As used herein, energetic refinement refers to the use of molecular mechanics simulation techniques, such as energy minimization or molecular dynamics, or other techniques, such as quantum-based approaches, to "adjust" the coordinates of a molecular structural model to bring it into a stable, low energy, conformation. In molecular mechanics simulations, the potential energy of a molecular system is represented as a function of its atomic coordinates along with a set of atomic parameters, called a [forcefield] force field. Energy minimization refers to a method wherein the coordinates of a molecular conformation are adjusted according to a target function [to result] that results in a lower energy conformation. Molecular dynamics refers to methods for simulating molecular motion by inputting kinetic energy into the molecular system corresponding to a specified temperature, and integrating the classical

equations of motion for the molecular system. During a molecular dynamics simulation, a system undergoes conformational changes so that different parts of its accessible phase space are explored.

Please amend the paragraphs beginning on page 14, line 12, through page 15, line 18, as follows:

As used herein, haplotype [refers] refers to two or more polymorphism located on a single DNA strand. Hence, haplotyping refers to identification of two or more polymorphisms on a single DNA strand. Haplotypes can be indicative of a phenotype.

As used herein, a parameter is any input data that will serve as a basis for sorting the database. These parameters will include phenotypic traits, medical histories, family histories and any other such information elicited from a subject or observed about the subject. A parameter may describe the subject, some historical or current environmental or social influence experienced by the subject, or a condition or environmental influence on someone related to the subject. [Paramaters] Parameters include, but are not limited to, any of those described herein, and known to those of skill in the art.

As used herein, computational phenotyping, refers to computer-based processes that assess the phenotype resulting from a particular genotype. The phenotype describes observables, such as, but are not limited to, the structure of the encoded protein, its functional morphological and structural attributes. In particular, as contemplated herein, the phenotype that is [assesed] assessed is the interaction of a protein with a particular [compounds] compound, particularly a drug. As exemplified herein, the method provides a means to select an effective drug for a particular [subjects] subject, particularly mammals, or class thereof.

As used herein, a database refers to a collection of data; in this case data relating to polymorphic variants. Hence a database contains the nucleic acid sequences encoding the variants, or a portion of the variant, such as a portion [contianing] containing the active site or targeted site. Additionally, the database may contain other information related to each entry, including but are not limited to, the corresponding 3-D structure of the encoded protein (or a portion thereof) and information regarding the source of each sequence. Some of the entries in a database may be identical, and for purposes herein, a database contains at least 2 different entries, typically far more than 2 entries. The number of entries depends upon the protein of interest and variety and number of polymorphisms that exist. Generally a database will have at least 10 different entries, typically more than 100, more than 500, more than 1000, more than 2000, 3000, 4000, 5000, 8000, 10,000, 50,000, 100,000 and greater. Databases herein containing 20,000 entries and more have been generated and are exemplified herein.

Please amend the paragraph beginning on page 22, line 28, through page 23, line 9, as follows:

It is shown herein that it is advantageous to use 3-D molecular structures in drug design rather than to consider primary sequence alone. For example, most drugs target proteins either in the afflicted organism or in a pathogen. Disease, drug action and toxicity are all manifested at the protein level. Although the nucleotide sequences of genetic polymorphisms might appear to be quite different, the resulting protein targets may have similar shapes and, therefore, the [protein] protein's biological function might be the same. Conversely, although genetic polymorphism sequences might appear similar, the resulting proteins may have critical differences in their 3-D structures that greatly affect biological activity. Thus, use of 3-D protein structure models in

such methods provide advantages not [heretofor] heretofore realized. Methods for generating 3-D structures are known to those of skill in the art and are also provided herein.

Please amend the paragraphs on page 24, lines 18-29, as follows:

The target gene is one that exhibits polymorphisms (i.e., sequence variations among a population) and the target protein is the product of a gene exhibiting genetic polymorphisms, or sequence variations, as described herein. Any gene or protein that exhibits polymorphisms is contemplated herein. In particular, genes that encode proteins, polypeptides, or oligopeptides that are targets for drug interaction are contemplated herein. The genetic polymorphisms can occur in the genes of pathogens (*e.g.* viruses, bacteriae, and fungi), parasites, plants, animals, and humans. As such, the sequence of a target protein can be obtained by the isolation and analysis of the gene or gene product in samples taken from pathogens, parasites, plants, animals, and humans, most preferably from humans.

Please amend the paragraphs beginning on page 29, line 22, through page 30, line 30, as follows:

Once the conserved regions of the model are assembled, *ab initio* loop prediction (Dudek *et al.* (1998) *J. Comp. Chem. 19*:548-573) indicated at 106A or *ab initio* secondary structure generation techniques of block 106B, techniques in which the alignments are adjusted using information on the secondary structure, functional residues, and disulfide bonds as described herein, can be used to complete the model (e.g. U.S. Patents Nos. 5,331,573; 5,579,250; and 5,612,895). This model, complete with loops, is then subjected to refinement procedures (block 110) based on molecular mechanics, molecular dynamics, and simulated annealing methods. Energetic refinement of the structure can be accomplished by performing molecular mechanics calculations using, for

example, an ECEPP type [forcefield] force field (Dudek *et al.* (1998) *J. Comp. Chem. 19*:548-573) or through molecular dynamics simulations using, for example, a modified AMBER type [forcefield] force field (Ramnarayan *et al.* (1990) *J. Chem. Phys. 92:7057-7076*. As known to those of skill in the art a modified AMBER (version 3.3) force field is a fully vectorized version of AMBER (3.0) with coordinate coupling, intra/inter decomposition, and the option to include the polarization energy as part of the total energy (see, *e.g.*, Weiner *et al.* (1986) *J. Comp. Chem. 7*:230-252). If necessary, the 3-D structures can be dynamically refined, for example, by using a simulated annealing protocol (*e.g.*,, 100 ps equilibration, 500 ps dynamics, up to 1000°K, 1 fs data collection).

The refinement process step 110 is used to offset problems that may arise when homology models are not built carefully or when they are built using fully automated methods. Problems that may arise include chain breaks (e.g. consecutive C^{α} atoms are farther apart than the optimum distance of 3.7 to 3.9 Å); distorted geometry (e.g. bond lengths and bond angles are too far from their optimal values); *cis*-peptide bonds (*e.g.*, incorrect isomerization of the peptide backbone in non-proline residues when it is not required); disallowed backbone and side-chain conformations (*e.g.*, dihedral angles do not satisfy the Ramachandran plot (see, Balasubramanian (1974) *Nature 266*:856-857) criteria for a fully favorable protein structure conformation); and misfolded loops (e.g. non-homologous loops are generated in unnatural conformations). The refinement procedure 110 removes distortions of covalent geometry by using energetic [methdods] methods, converts disallowed backbone and side-chain conformations into allowed ones using simulated annealing methods, conserves protein core structure and secondary structural elements built by homology, and

rebuilds unnatural loop constructions (Dudek *et al.* (1998) *J. Comp. Chem. 19*:548-573).

Please amend the paragraph beginning on page 32, line 15, through page 33, line 2, as follows:

Next, at block 214, the 3-D structural models for all variants are generated. A refinement process is then completed at block 216 for the structural models. As noted above in connection with FIG. 1, the process involves subjecting each model, complete with loops, to refinement procedures based on molecular mechanics, molecular dynamics, and simulated annealing methods. As before, the energetic refinement of the structure can be accomplished by performing molecular mechanics calculations using an ECEPP type [forcefield] force field (Dudek *et al.* (1998) *J. Comp. Chem. 19*:548-573), or through molecular dynamics simulations using, for example, a modified AMBER type [forcefield] force field (Ramnarayan *et al.* (1990) *J. Chem. Phys. 92*:7057-7076), where a modified AMBER (version 3.3) force field is a fully vectorized version of AMBER (3.0) with coordinate coupling, intra/inter decomposition, and the option to include the polarization energy as part of the total energy (Weiner et al. (1986), J. Comp. Chem. 7:230-252). If necessary, the 3-D structures can be dynamically refined, for example, by using a simulated annealing protocol (*e.g.*,, 100 ps equilibration, 500 ps dynamics, up to 1000°K, 1 fs data collection).

Please amend the paragraph on page 34, lines 1-9, as follows:

At block [328] 228, once the models are determined to be satisfactory, drug molecules are docked with the structural variant models. Next, at block 330, the free energy of binding is evaluated with the potential drugs under study for each structural variant model. At block 332, the total free energy of binding is decomposed, based on the interacting residue in the protein active

site. Lastly, at block 334, the free energy of binding is correlated with patient data, if the data is available. Thus, the 3-D structural data is employed in drug design. Details of using such structural data in drug design are described further below.

Please amend the paragraph on page 34, lines 11-15, as follows:

The crystal structure of any protein can be determined empirically and the resulting coordinates used as the basis for [determing] determining structures of variants. Such structures are often known (see, *e.g.*, Kohlstaedt *et al.* (1992) *Science 256*:1773-1790 for a crystal structure of HIV-1 RT bound to a ligand).

Please amend the paragraph beginning on page 38, line 13, through page 39, line 8, as follows:

New potential drug candidates can be designed by identifying potential small molecule drugs that can bind to a particular structural variant. This is accomplished, for example, by methods including, but are not limited to, methods for electronic screening of small molecule databases as described herein, methods involving modifying the functional groups of existing drugs *in silico*, methods of *de novo* ligand design. Methods for computationally [desiging] designing drugs are known to those of skill in the art and include, but are not limited to, DOCK (Kuntz *et al.* (1982) "A Geometric Approach to Macromolecule-Ligand Interactions", J. Mol. Biol., 161:269-288; available from University of Ca, San Francisco); and AUTODOCK (see, Goodsell *et al.* (1990) "Automated Docking of Substrates to Proteins by Simulated Annealing", Proteins: Structure, Function, and Genetics, 8, pp. 195-202; available from Scripps Research Institute, La Jolla); GRID (Oxford University, Oxford, UK); CAVEAT (UC Berkeley, Ca), LEGEND (Molecular Simulations, Inc., San Diego, CA); LUDI (Molecular Simulations, Inc., San Diego, CA); HOOK (Molecular Simulations, Inc., San Diego, CA); CLIX (CSIRO, Australia); GROW (Upjohn

Laboratories, Kalamazoo); others including HINT, LUDI, NEWLEAD, HOOK, PRO-LIGAND and CONCERTS (see, M. Murcko, "An Introduction to De Novo Ligand Design" in Practical Application of Computer-Aided Drug Design, Charifson, Ed., Marcel Dekker, NY, pp 305-354), methods based on QSAR (quantitative structure-activity relationships, *QSAR and Drug Design: New Developments and Applications*, Fugita, Ed., (1995) Elsevier, pp 3-81; 3D QSAR in Drug Design, Kubinyi, Ed., (1993) Escom, Leiden), and other methods known to those of skill in the art for determining molecules that have optimal binding interactions with a selected target.

Please amend the paragraph beginning on page 39, line 15, through page 40, line 4, as follows:

After the computational docking step, the free energy of binding of the docked complex is calculated, and the total free [enegy] <u>energy</u> of binding is decomposed based on the interacting residues in the protein active site or sites deemed [improtant] <u>important</u> for protein activity. Analyses of the binding energies are needed to identify drug candidates. If needed or desired, the free energy of binding of different drugs or potential drugs to each structural variant model can be calculated by [substracting] <u>subtracting</u> the free energy of the non-interacting protein and drug from the free energy of the protein-drug complex. The total free energy of binding is decomposed into its various thermodynamic components, e.g. enthalpic and entropic components, based on the interacting residues in the protein active site in a solvated model to characterize the structural and thermodynamic features in the mode of drug binding and to determine the contribution of the solvent (see, *e.g.*, Wang *et al.* (1996) *J. Am. Chem. Soc. 118*:995-1001; Wang *et al.* (1995) *J. Mol. Biol. 253*:473-492; Ortiz *et al.* (1995) *J. Med. Chem. 38*:2681-2691, which describes a computational method for deducing QSARs from ligand-

macromolecule complexes). Following the computational drug design protocol described herein, any potential new drugs that are identified can be synthesized in, for example, industry or academia, and subjected to further biological testing, such as *in vitro* studies or pre-clinical and clinical *in vivo* testing.

Please amend the paragraph on page 44, lines 6-16, as follows:

If common structural features are observed over a range of protein targets that are derived from genetic polymorphisms, these common features may be used to design a drug that is effective with a variety of genetic polymorphisms and thus many patients. The retention of certain common structural features over a large number of genetic polymorphisms suggests that those features may not be [mutatable] mutable because the conserved structure may be essential to protein function, *e.g.*, to the viability of an infectious organism or virus. Such conserved structural elements are prime targets for structure-based drug design, *e.g.*, anti-infective or antibiotic drug design, and can lead to highly effective therapies.

Please amend the paragraph beginning on page 44, line 28, through page 45, line 14, as follows:

In comparing sets of related protein structures, such as those with the same biological function or those resulting from genetic polymorphisms, certain parts of the structural framework are often found to be conserved, while other parts vary among the proteins. Mutations that occur in the conserved regions of the structure can have significant effects biological activity. For example, in viruses, the conserved features can be essential to protein function and, thus, to the viability of the infectious organism or virus. Identifying the conserved structural features over a range of structures often gives insight into which structural features are necessary for biological activity and are therefore [non-mutatable] non-mutable. By analyzing a number of structural variants derived

from genetic polymorphisms that exhibit drug resistance, it is possible to identify or design drugs that interact best with the common structural features in all of the variants. Using these features in structure-based drug design studies leads to the identification of drugs that retain biological activity despite multiple mutations, or polymorphisms, and could help to overcome the problem of drug resistance.

Please amend the paragraph on page 51, lines 10-29, as follows:

A database is preferably interfaced to a molecular graphics package that includes 3-D visualization and structural analysis tools, to analyze similarities and variations in the protein structural variant models (see, copending U.S. application Serial No. 09/531,995, which is published as International PCT application No. WO 00/57309, and is a continuation-in-part of U.S. application Serial No. 09/272,814, filed March 19, 1999). Briefly, International PCT application No. WO 00/57309 provides a database and interface for access to 3-D molecular structures and associated properties, which can be used to facilitate the design of potential new therapeutics. The interface also provides access to other structure-based drug discovery tools and to other databases, such as databases of chemical structures, including fine chemical or combinatorial libraries, for use in structure-focused high-throughput screening, as well as to a host of public domain databases and bioinformatics sites. The interface also provides access to other structure-based drug discovery tools and to other databases, such as databases of chemical structures, including fine chemical or combinatorial libraries, for use in structure-focused high-throughput screening, as well as to a host of public domain databases and bioinformatics sites. This interface can be modified as needed to adapt for use with a [paritcular] <u>particular</u> database.

Please amend the paragraph on page 54, lines 20-29, as follows:

Databases containing data representative of the 3-D structure of structural variants encoded by a selected gene or genes or the 3-D structure of other polymorphic variants are provided. The selected genes can be genes of drug targets, such as receptors, and genes of infectious agents, such as the HIV protease or reverse transcriptase. Exemplary databases are presented in Example 5 which describes the construction, interface, use and [appliations] applications of HIV PR and RT databases. These databases may be stored on any suitable medium and used in any suitable computer system. Systems and methods for generating, storing and processing databases are well known.

Please amend the paragraph on page 69, lines 20-25, as follows:

To modify these compounds, important pharmacophore features on the surface of the receptor that are critical for binding of the compounds were identified. These features include a hydrophobic belt, a hydrophilic belt and 3 hydrogen bond donor sites. A few [of] potential hydrogen bonding sites, which are not used by the current compounds, were also derived, and can be used for designing more potent binders.

Please amend the paragraph beginning on page 76, line 30, through page 77, line 2, as follows:

Computational or *in silico* phenotyping is performed to assess phenotypic properties of a protein. This example [demosntrates] demonstrates application of this method to HIV-1 protease and reverse transcriptase to test whether the efficacy of various protease inhibitors for an HIV patient.

IN THE CLAIMS:

Please amend claims 1, 3, 4, 15, 24, 25, 45, 48, 49, and 87 as follows:

1. (Amended) A computer-based method of drug design based on genetic polymorphisms, comprising:

identifying target proteins that are the product of a gene exhibiting genetic polymorphisms;

obtaining more than one amino acid sequence of the target proteins that are the product of a gene exhibiting genetic polymorphisms, wherein the sequences represent different genetic polymorphisms;

[generating] determining 3-dimensional (3-D) protein structural variant models [from]for the [sequences] target proteins that are the product of a gene exhibiting genetic polymorphisms; and

based upon the structures of the 3-D models of the target proteins that are the product of a gene exhibiting genetic polymorphisms, designing drug candidates, modifying existing drugs, identifying potential drug candidates or identifying modifications of existing drugs based on predicted intermolecular interactions of the drug candidates or modified drugs with the structural variants of the target proteins.

3. (Amended) The method of claim 2, wherein the binding interactions are determined by:

calculating the free energy of binding between the protein structural variant model and the docked molecule; and

decomposing the total free energy of binding based on the interacting residues in the protein active site.

4. (Amended) The method of claim 1, wherein:

after the protein structural variant models derived from a particular genetic polymorphism are generated, selected model structures are analyzed to determine common structural features that are conserved throughout the selected models, wherein

the conserved structural features are used as a basis for structure-based drug design studies.

15. (Amended) The method of claim 1, wherein:

after [generating] determining the 3-D protein structural variant models, the method comprises:

computationally docking drug molecules with the target protein models; and

energetically refining the docked complexes; and

wherein the candidate drugs are specific for a protein with a selected polymorphism or specifically interact with all proteins exhibiting a polymorphism.

24. (Twice Amended) The method of claim [13]15, wherein the structural variant models are stored in a relational database, comprising:

3-D molecular coordinates for the structural variants;

a molecular graphics interface for 3-D molecular structure visualization; [and]

computer functionality for protein sequence and structural analysis; and

database searching tools.

25. (Amended) The method of claim [13]24, wherein the database further comprises observed clinical data associated with the genetic polymorphisms, subject medical history and subject history.

45. (Amended) The method of claim 1, wherein the target protein is an enzyme.

48. (Amended) The method of claim 45, wherein the target protein is a protein expressed by an infectious agent.

49. (Twice Amended) The method of claim 45, wherein the target protein is an enzyme expressed by an infectious agent.

87. (Amended) The method of claim [1]12, wherein the selected subpopulation is a human [patient] subject subpopulation.